

PATENT APPLICATION

VOICE MENU SYSTEM

Inventor:

Alexander B. Beaman
4024 Ross Park Court
San Jose, CA 95118-1758

Assignee:

Apple Computer, Inc.
1 Infinite Loop
Cupertino, CA 95014

Prepared by:

BEYER, WEAVER & THOMAS, LLP
P.O. Box 778
Berkeley, CA 94704-0778

VOICE MENU SYSTEM

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates to media players and, more particularly, to navigating menus of media players.

Description of the Related Art

[0002] The ability of computers to be able to share information is of utmost importance in the information age. Networks are the mechanism by which computers are able to communicate with one another. Generally, devices that provide resources are called servers and devices that utilize those resources are called clients. Depending upon the type of network, a device might be dedicated to one type of task or might act as both a client and a server, depending upon whether it is giving or requesting resources.

[0003] Increasingly, the types of resources that people want to share are often entertainment-related. Specifically, music, movies, pictures, and print are all types of entertainment-related media that someone might want to access from across a network. For example, although a music library may reside on a desktop computer, the media owner may want to listen to the music on a portable media player.

[0004] In order to achieve portability, many portable media players use minimalist displays that allow the user access to the music via simple graphical user interfaces. The displays are not always well-lit, and may not be navigable in the dark. Also, the user may be in certain situations (e.g., driving a car) where it is not convenient or appropriate to look at the display, or may have a physical disability that makes

visually navigating the menu impossible. Additionally, many people may simply find the displays too small and inconvenient to use on a regular basis.

[0005] Although the described technologies work well in many applications, there are continuing efforts to further improve the user experience.

SUMMARY OF THE INVENTION

- [0006] The present invention provides a method for providing an audio menu. First, text strings are provided on a server, each text string being capable of representing a menu choice. Next, audio files are generated, each audio file representing a voiced name of one of the text strings, and each audio file is associated with its text string. The server then delivers both the audio files and the associations to a client.
- [0007] A menu is subsequently presented on the client that includes menu choices represented by the text strings, the menu choices being capable of being highlighted or selected. The audio files are played on the client when their associated menu choices are highlighted.
- [0008] In another aspect of the invention, a server that includes a processor, memory and a network interface is provided. The server's processor is operable to perform instructions including providing a text string that represents a menu component, whereby the menu component is one of several options that can be selected from a menu on a client device. The server's processor is also operable to perform other instructions such as generating an audio file that is an audio representation of the menu component and delivering the audio files to a client device.
- [0009] In yet another aspect of the invention, a client device that includes a processor, memory and a network interface is provided. The client's processor is operable to perform instructions including allowing it to receive an audio file from a server that is an audio representation of a menu component, whereby the menu component is one of several options that can be selected from a menu. The client's

processor is also operable to perform instructions that include allowing it to update the menu to include the menu component and playing the audio file when the menu component is highlighted.

[0010] In yet another aspect of the invention, a media management system is provided. The media management system includes a media database, media collection records, media records, a voiced names database and string association records. The media database stores media files. The media collection records include data relating to groupings of the media files. The media records include metadata relating to the media files. The voiced names database stores audio files. The string association records associate the audio files with data from the media collection records and metadata from the media records.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention may best be understood by reference to the following description taken in conjunction with the accompanying drawings in which:

FIG. 1 is a block diagram illustrating an exemplary environment in which the present invention may be implemented;

FIG. 2 is a block diagram illustrating an organizational structure of a media management system according to one embodiment of the invention;

FIG. 3 is a flow chart illustrating the general steps that can be used in connection with one embodiment of the invention;

FIG. 4 is a flow chart illustrating one possible method of generating voiced names according to one embodiment of the invention as required in FIG. 3;

FIG. 5 is a flow chart illustrating the steps that can be performed in activating the audible menu option in a client device according to one embodiment of the invention;

FIG. 6 is a flow chart illustrating the steps that can be performed during menu navigation according to one embodiment of the invention; and

FIG. 7 is a diagram illustrating an exemplary computing device in which various embodiments of the invention may be implemented.

It is to be understood that, in the drawings, like reference numerals designate like structural elements. Also, it is understood that the depictions in the figures are not necessarily to scale.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0011] In the following description, numerous specific details are set forth to provide a thorough understanding of the present invention. However, it will be obvious to one skilled in the art that the present invention may be practiced without some or all of these specific details. In other instances, well known process steps have not been described in detail in order to avoid unnecessarily obscuring the present invention.

[0012] The present invention generally allows for updateable audio menus. Although a device might have some pre-packaged menu components, other menu components are received from a server. For example, a music player might come with some pre-installed menu components (e.g., a top menu level of “Playlists,” “Songs,” “Artists,” “Settings,” and “About”) but allow other menu components to be added to the various menu choices (e.g., a user-added top level menu of “Genre” or the second level menu listing of available playlists, songs, and artists). Each menu component, regardless of whether it is original or received from the server, has an associated voiced name. When the user highlights a menu choice, the voiced name is played. The user then has the option of selecting the menu choice or scrolling to a new menu choice. In this way, a user could navigate the menu without having to see the visual display.

[0013] FIG. 1 is a block diagram illustrating an exemplary environment in which the present invention may be implemented. A network 105 couples a server 110 to various clients 115, 120, 125, and 130. The network 105 can generally be a data network, such as a LAN, WAN or the Internet. The server 110 may or may not be a dedicated device. In the example shown in FIG. 1, the server 110 is a general purpose

computer. The various clients 115, 120, 125, and 130 can be thick or thin clients, with varying levels of processing power. Clients may include portable computers 115, desktop computers 120, specialized devices such as iPodsTM 125 available from Apple Computer, Inc. of Cupertino, California, or even network-aware audio/video components 130 that are designed to work across the network 105. Certain devices, such as the iPod 125, might directly connect to the server 110 via FireWire, USB, or some other external bus that allows the client 125 and the server 110 to be more directly networked together.

[0014] FIG. 2 is a block diagram illustrating an organizational structure of a media management system 200 according to one embodiment of the invention. The media management system 200 is the computer program that allows the user to both organize and access digital media. For simplicity, the following discussion will assume the digital media is limited to music. It should, however, be appreciated that any reference to “songs” or “music” could be generalized to any form of digital media, which can include sound files, picture data, movies, text files or any other types of media that can be digitally stored on a computer. Similarly, any reference to “playlists” can be generalized to media collections, including collections of mixed digital media.

[0015] Although the server 110 and the clients 115, 120, 125, and 130 might each have different versions of the media management system 200 that are specially tailored to the specific functionality required from those devices, the basic components of the media management system 200 are similar. Specifically, the media management system 200 can include a media manager 205, a music database

210, and a voiced names database 215. The media manager 205 manages the databases 210 and 215.

[0016] The music database 210 has a number of song records 220 and playlist records 225 that are used to classify, identify and/or describe media (i.e., media items) in the music database 210. The song records 220 contain metadata about each media item available in the database 210. The metadata might include, for example, the names of songs, the artist, the album, the size of the song, the format of the song, and any other appropriate information. Of course, the type of information may depend on the type of media. A video file might additionally have director and producer fields, but may not use the album field.

[0017] The playlist records 225 contain information about each playlist available in the music database 210. Further, the information for a given playlist can include identifying information for each of the songs within the playlist. Playlists are collections of media that may or may not be in any particular order. Users may choose to combine media by genre, mood, artists, audience, or any other meaningful arrangement.

[0018] Some of the information contained in the various records 220, 225, and 230 are used as menu components. For example, a top level of menu components might permit the user to navigate through “Songs,” “Artists,” or “Playlists.” These classifications could either be pre-packaged with the media management system 200, or modified by the user if the media management system 200 permits modification. The user would then be able to navigate to a specific media through several different routes.

[0019] For example, if the user wanted to access the song “Little Angel of Mine” through the “Songs” menu component, the user would scroll through the top level choices until the “Songs” menu component was highlighted. Once highlighted, the user would select “Songs” and be presented with a second level listing of menu components. This second level listing could simply be an alphabetical list of all songs available to the user, with each song being a second level menu component. Typically, none of these second level menu components would be pre-packaged, and they would all rely upon the user’s particular music preferences. The user would scroll through the songs until “Little Angel of Mine” was highlighted, and then select that menu component to play the song.

[0020] Alternatively, if the user wanted to access the song through “Artists,” the user would scroll through the top level of menu components until “Artists” was highlighted, and then select “Artists” in order to be presented with the second level of menu components. The user would then scroll through an alphabetical listing of artists until the group “No Secrets” was highlighted. Selecting the “No Secrets” second level menu component would then take the user to a third level of menu components that listed all the songs performed by the group No Secrets. The song “Little Angel of Mine” would then be among the third level menu components.

[0021] Yet another alternative method of navigating to the sound would be to access the song through the user-defined playlists. Selecting the top level menu component “Playlists” would bring the user to a second level listing of all the playlists the user had previously created. The song “Little Angel of Mine” might be listed under several different playlists. For example, either the “Stuart Little 2 Soundtrack” or “Songs Written by Orrin Hatch” playlist might contain the song. Selecting either

of these second level menu components would bring the user to a third level listing of songs in that playlist.

[0022] Each of the described menu components is taken directly from the records 220 and 225. Associated with each menu component is an audio representation of that menu component. In the previous example, “Songs,” “Artists,” “Playlists,” “No Secrets,” “Stuart Little 2 Soundtrack,” “Songs Written by Orrin Hatch,” and “Little Angel of Mine” would all require associated vocalizations to allow a user to navigate the menus without any visual cues.

[0023] One mechanism for maintaining the vocalizations is the voiced names database 215. The voiced names database 215 contains an audio file for each vocalization and a number of records 230 that maintain the associations between the audio files and their corresponding menu components. Although alternative mechanisms are possible (e.g., embedding the vocalizations in the song records 220 and playlist records 225, thereby eliminating the need for a voiced names database 215), using a separate voiced names database 215 allows a single vocalization to be used independently of how the user navigates to a particular menu component.

[0024] FIG. 3 is a flow chart illustrating the general steps that can be performed in connection with one embodiment of the invention. At 305 a text string representing a new menu component is introduced to the server 110. This introduction can occur via a user manually entering a new entry, such as a new playlist, or the introduction can occur automatically, such as with the purchase of a new song file that comes packaged with a song record 215.

[0025] At 310, if necessary, an audio file is generated that is the voiced name of the menu component. Generation of a voiced name might not be necessary if a

purchased song included a voiced name, or if the voiced name already exists in the voiced name database 215. For example, if the user already had a voiced name for “The Beatles” there would not be a need to create a duplicate voiced name every time a new Beatles song was added to the music database 210.

[0026] FIG. 4 is a flow chart illustrating the detailed steps involved in generating voiced names according to one embodiment of the invention. At 405, the media management system 200 receives a trigger to create a voiced name. Typically, the trigger will be the creation of a new menu component through the introduction of a new song record 220 or a new playlist record 225. However, if the voiced name option was previously turned off, turning the option on for the first time would generate a trigger that informed the media management system 200 that voiced names are required.

[0027] Once a trigger is generated, at 410 the media management system 200 determines whether a voiced name for the particular string already exists. If a voiced name does not exist, the server 110 could use standard text-to-speech tools to generate audio files at 415. Preferably, the files will additionally be compressed to save space. One popular codec that both encodes and compresses speech is Qualcomm PureVoice, available from Qualcomm, Inc. of San Diego, California.

[0028] Once an audio file is created, the server 110 could optionally play the voiced name for the user at 420 so that the user can hear the audio file. At 425 the user can be given the option of approving or rejecting the vocalization. If the user approves of the vocalization, then the media management system 200 will create the appropriate string association record 230 at 430 so that the audio file is associated with the appropriate menu component.

[0029] Otherwise, if the user does not approve of the vocalization at 425, the user might be given the option to change the text that the text-to-speech tool uses to create a voiced name at 435. The text that the user inputs could optionally be independent of the menu component, allowing the user to sound out the menu component phonically without altering the actual text that is used in the records 220 and 225, therefore allowing the menu component to be both spelled correctly and pronounced correctly. The new vocalization could then be played back to the user at 420, giving the user the choice to approve of the new vocalization.

[0030] Alternatively, if the user chooses not to change the text at 435, the media management system 200 might allow the user to record his or her own vocalization at 440 or possibly provide another audio file. The user's own voice could then be used to later navigate the menus.

[0031] Referring back to FIG. 3, after an audio file of a voiced name is created at 310, the server 110 delivers any new files to a client device 115, 120, 125 or 130 at 315. Typically, the contents of the voiced names database 215 and the string association records 230 will be delivered when the user is downloading the music database 210 and their associated records 220 and 225 from the server 110 to the client device 115, 120, 125 or 130. However, there is no reason why the voiced names database 215 and association records 230 could not be delivered independently of the music database 210 and its records 220 and 225.

[0032] At 320, the client device 115, 120, 125 or 130 receives the audio files, along with any new menu components as appropriate. Once received, the menus on the client's media management system 200 are updated at 325 to reflect any changes. Then, at 330, whenever the user highlights any menu component, the appropriate

audio file is played back to the user, allowing the user to navigate the menu by audio cues.

[0033] Typically, the media management system 200 will give the user the option of turning on or off the audible menus. FIG. 5 is a flow chart illustrating the steps that can be performed in setting the audible menu option according to one embodiment of the invention. At 505, the user can optionally select a language option. The language option allows the pre-packaged menu components to be presented in other languages. For example, the “Songs” menu component would be presented to the user as “Canciones” in Spanish, “Chansons” in French and “Canzoni” in Italian. Additionally, the English-version voiced name would no longer be appropriate, and could be substituted with an appropriate foreign language vocalization. The foreign language vocalization could either be pre-packaged in the media management system 200, or could require downloading from the server 110. Typically, once language options are set, they will not be changed.

[0034] At 510 the user activates the audible menu feature. While this might cause the client device 115, 120, 125, or 130 to use pre-defined settings, it is also possible to present the user with various customization options. For example, at 515, the user can choose to play music while browsing the menu. Once the user selects a song to be played, the user might want to queue up another song while listening to his or her first selection. Accordingly, the user can be given the option of allowing the voiced names to be presented while the first selected song is playing. If the user does not want the music to play during menu navigation, the system can be set up to either pause or mute the music at 520.

[0035] If the user wants to hear music while navigating the menus, at 525 the user can be given the option of mixing the music with the voiced names. Mixing is simply done by playing the audio file over the currently playing song. If mixing is desired, the mixing option is set at 530. If mixing is not desired, but the user still wants the music to play while navigating the menus, the media management system 200 could allow the music to play in one channel (either the left or right speaker) and play the voiced name in the remaining channel by setting the single-channel option at 535. Therefore, when the user is wearing headphones, the voiced names would be presented in one ear while the music would be playing without interruption in the other ear. Additionally, even if the user selected the mixing option at 530 or the pause music option at 520, there is no reason why the user could not also select the voiced names to be output in a single channel at 540.

[0036] Once all the audible menu features are set, the client device 115, 120, 125, or 130 is ready to use the voiced names during menu navigation. FIG. 6 is a flow chart illustrating the steps that can be performed during menu navigation according to one embodiment of the invention.

[0037] At 605 the menu is activated. While activation may not be required if the menu is always active, some client devices 115, 120, 125, or 130 might hibernate the menu after periods of inactivity. Typically, the menu is taken out of hibernation by pressing a navigation control. The navigation controls can include dials, buttons, touch screens, or any other convenient input mechanism. The navigation controls could be present on the client device 115, 120, 125, or 130, or through a remote control. It should be appreciated that many remote controls do not have any visual

displays, making menu navigation inconvenient if a visual display on the client device 115, 120, 125 or 130 must be used.

[0038] Once active, the media management system 200 optionally determines whether the menu component has been highlighted for a sufficient period of time at 610. It might be very distracting for the user to scroll through the menu components and hear the voiced names of each menu component start, only to be cut off by the voiced name of the next menu component, which is then cut off by the voiced name of the next menu component. Preferably, the media management system 200 will have a slight delay so that the user could rapidly scroll through the various options without such a distraction. At 615 the media management system 200 waits until the user stops scrolling through the menu components, and pauses for enough time on a single menu component to permit a voiced name to be played at 620. The period of time does not have to be long, and will typically be no more than a few seconds, and may even be some fraction of a second.

[0039] At 625 the user then has the option to navigate to a new menu component and restart the process. Navigation can either be done through scrolling, or, if the currently highlighted menu component leads to another level of menus, through selection of the current menu component. Alternatively, the process can end if the user simply stops navigating the menus, or makes a menu component selection that does not lead to more menu choices (e.g., playing a song).

[0040] Generally, the techniques of the present invention may be implemented in software and/or hardware. For example, they can be implemented in an operating system, in a separate user process, in a library package bound into applications, or on a specially constructed machine. In a specific embodiment of this invention, the

technique of the present invention is implemented in software such as an operating system and/or in an application program running on the operating system.

[0041] A software or software/hardware hybrid implementation of the techniques of this invention may be implemented on a general-purpose programmable machine selectively activated or reconfigured by a computer program stored in memory. In an alternative embodiment, the techniques of this invention may be implemented on a general-purpose network host machine such as a personal computer, workstation or server. Further, the invention may be at least partially implemented on a general-purpose computing device.

[0042] Referring now to FIG. 7, a computing device 700 suitable for implementing the techniques of the present invention includes a master central processing unit (CPU) 705, interfaces 710, memory 715 and a bus 720. When acting under the control of appropriate software or firmware, the CPU 705 may be responsible for implementing specific functions associated with the functions of a desired computing device. The CPU 705 preferably accomplishes all these functions under the control of software including an operating system (e.g., Mac OS X), and any appropriate applications software (e.g., iTunes).

[0043] CPU 705 may include one or more processors such as those from the Motorola family of microprocessors or the MIPS family of microprocessors. In an alternative embodiment, the processor is specially designed hardware for controlling the operations of computing device 700.

[0044] The interfaces 710 are typically provided as interface cards. Generally, they control the sending and receiving of data packets over the network and sometimes support other peripherals used with the computing device 700. Among the

interfaces that may be provided are Ethernet interfaces, frame relay interfaces, cable interfaces, DSL interfaces, token ring interfaces, and the like. In addition, various very high-speed interfaces may be provided such as fast Ethernet interfaces, Gigabit Ethernet interfaces, ATM interfaces, HSSI interfaces, POS interfaces, FDDI interfaces, ASI interfaces, DHEI interfaces, Firewire interfaces, USB interfaces and the like. Generally, these interfaces may include ports appropriate for communication with the appropriate media. In some cases, they may also include an independent processor and, in some instances, volatile RAM.

[0045] Regardless of computing device's configuration, it may employ one or more memories or memory modules (such as, for example, the memory 715) configured to store data, program instructions and/or other information relating to the functionality of the techniques described herein. The program instructions may control the operation of an operating system and/or one or more applications, for example.

[0046] Because such information and program instructions may be employed to implement the systems/methods described herein, the present invention relates to machine (e.g., computer) readable media that include program instructions, state information, etc. for performing various operations described herein. Examples of machine-readable media include, but are not limited to, magnetic media such as hard disks, floppy disks, and magnetic tape; optical media such as CD-ROM disks; magneto-optical media such as floptical disks; and hardware devices that are specially configured to store program instructions, such as read-only memory devices (ROM) and random access memory (RAM). The invention may also be embodied in a carrier wave traveling over an appropriate medium such as airwaves, optical lines, electric

lines, etc. Examples of program instructions include both machine code, such as produced by a compiler, and higher level code that may be executed by the computer (e.g., using an interpreter).

[0047] Although illustrative embodiments and applications of this invention are shown and described herein, many variations and modifications are possible which remain within the concept, scope, and spirit of the invention, and these variations would become clear to those of ordinary skill in the art after perusal of this application. For example, the terms “scroll” and “highlight,” when used in the context of menus, are not limited to their literal interpretations. Menu choices can be “scrolled” on a single line, with one menu component replacing the last menu component. Similarly, a menu choice can be “highlighted” even if it is italicized, bolded, or listed with a bullet. Accordingly, the present embodiments are to be considered as illustrative and not restrictive, and the invention is not to be limited to the details given herein, but may be modified within the scope and equivalents of the appended claims.